land

MDPI

Article

The Scale Effect of Street View Images and Urban Vitality Is Consistent with a Gaussian Function Distribution

Jie Yin [1,2,*], Ran Chen [1,2], Ruixiang Zhang [1,2], Xiang Li [1,2] and Yuan Fang [2,*]

[1] School of Urban Planning and Design, Shenzhen Graduate School, Peking University, Shenzhen 518055, China; chenran_alyssa@stu.pku.edu.cn (R.C.); 2401212697@stu.pku.edu.cn (R.Z.); lixiang1999@stu.pku.edu.cn (X.L.)

[2] Key Laboratory of Earth Surface System and Human-Earth Relations of Ministry of Natural Resources of China, Shenzhen Graduate School, Peking University, Shenzhen 518055, China

* Correspondence: yinj@pkusz.edu.cn (J.Y.); fangyuan@stu.pku.edu.cn (Y.F.)

Abstract: Street view images are often used to assess the impact of the built environment on urban vitality from an eye-level perspective. However, the influence of analyzing scale on the accuracy of assessment results is often ignored. To find the appropriate scale, we need to quantify the scale effect of street view images and urban vitality. Therefore, in this study, Shenzhen is selected as the study area. Urban vitality is characterized by Weibo check-in records. VGG19 is employed to learn the features of street view images and performs regression analysis of the image features and Weibo check-in records in different sizes of grids (100 m, 150 m, . . ., 1000 m). Finally, we fit the scale effect of their correlation via a function. We find that as the scale increases, the correlation between street view images and urban vitality tends to increase and then decrease, which is consistent with the distribution law of a Gaussian function. This study provides a basis for selecting appropriate scales of the correlation between street view images and urban vitality.

Keywords: street view image; scale effect; urban vitality; Gaussian function; VGG19

check for updates

Academic Editors: Seung-Nam Kim, Gi-Hyoug Cho and Hee-Chul Kim

Received: 19 December 2024
Revised: 4 February 2025
Accepted: 13 February 2025
Published: 17 February 2025

Citation: Yin, J.; Chen, R.; Zhang, R.; Li, X.; Fang, Y. The Scale Effect of Street View Images and Urban Vitality Is Consistent with a Gaussian Function Distribution. *Land* **2025**, *14*, 415. https://doi.org/10.3390/land14020415

1. Introduction

Street view images have attracted widespread attention because they can be used to assess built environment features from an eye-level perspective [1,2]. Street view images are massive image data captured along streets. Previous studies often quantify built environment features from an eye-level perspective by classifying, detecting, and segmenting images [3]. However, the scale effect is not considered when these images are used [4]. The scale effect refers to variation in results obtained by applying the same statistical analysis method at different scales [5]. When utilizing street view images, the grid size is typically defined as the spatial scale [6], and the built environment features are assessed through the visual appearance characteristics extracted from these images within the grid [7,8]. The spatial variations can be seen in the visual appearance characteristics. These variations change with the analyzing scales [4], leading to the modifable area unit problem (MAUP) [9]. Therefore, it is necessary to explore the appropriate scale of street view images.

The appropriate scale of street view images needs to be discerned by correlation studies featuring influencing factors. A large number of studies have used street view images to evaluate the correlation between the built environment from an eye-level perspective and its influencing factors, but there is a lack of discussion on the reasons for the selection of the scale of analysis [8,10]. Among the influencing factors, the significant correlation between street view images and urban vitality has been widely mentioned [10–13]. Urban vitality is the comprehensive activity level in the economic, social, and cultural aspects of a city under the influence of human activities. It is essential for assessing the city's development potential and competitiveness [14]. Urban vitality is often characterized by the spatial concentration of human activities. Location-Based Services (LBSs) such as social media [15], GPS [16], and POI [17] are widely used to characterize urban vitality. Understanding the impact of street view images on urban vitality can help urban planners optimize the design of street façades to improve urban vitality. However, previous studies on the correlation between street view images and urban vitality often rely on empirical selection using a single scale [8,18]. Different studies obtain different results at various scales, reducing the comparability of findings. For example, building continuity derived from street view images was found to consistently correlate with urban vitality at the 500 m grid scale [8]. In contrast, a study at the community scale revealed that the impact of building continuity was less pronounced, as expected [19]. It suggests that the correlation between street view images and urban vitality can be significantly affected with scale changes. It makes some study conclusions not credible. Therefore, it is necessary to explore the scale effect of street view images and urban vitality. With a focus on people-oriented urban design, a deeper understanding of how the scale effect impacts the correlation between street view images and urban vitality can provide a reference for designing unified street façades in zoning plans.

Previous studies have explored the scale effects of the built environment and urban vitality, suggesting that environmental planimetric features such as intersection density and building density are more relevant to urban vitality at smaller scales [20]. However, the selection of different research areas will lead to various optimal scales. For example, examining the related mechanism of the built environment and urban vitality and scale changes in Nanjing, China, shows that the most relevant spatial scale is 100 m [18]. In contrast in a study in Tampere, Finland, it was found that the most relevant spatial scales are 200 m and 300 m [20]. There are two main limitations in previous studies of the built environment and urban vitality. On the one hand, the discussion of optimal scales is only applicable to certain cities. Without quantifying the scale effect, it is difficult to generalize the results to other study areas, and further work is needed to identify universal laws by fitting the scale effect with a function. On the other hand, previous studies have mainly focused on the scale effect of environmental planimetric features and urban vitality, while built environment façade characterization based on street view images has not received much attention.

To quantify the scale effect, previous studies often adopt function fitting [21]. By analyzing the characteristics of the numerical distribution of the correlation index with scale changes, the fitting function is determined. Based on the fitting effect, the explainability of the phenomenon is judged. Fitting the scale effect between street view images and urban vitality via a function can reveal the general law of the correlation between built environment façade features and urban vitality through scale changes. Although cities have different optimal scales for correlations between street view images and urban vitality, the law of scale-dependency should have common mathematical properties. By describing the complex scale effect with a simple mathematical model, the study can be more generalizable.

To realize the above research purpose, Shenzhen is divided into different sizes of grids, such as 100 m, 150 m, . . ., 1000 m, to reflect scale changes. First, to recognize the façade features of the built environment, we perform feature learning for street view images based on deep learning. Second, the regression analysis of urban vitality characterized by Weibo check-in records is performed at different scales. Finally, we analyze the numerical distribution of the strength of correlation between street view images and urban vitality, and explore the scale effect of correlation by fitting a function. By quantifying the scale effect, this process not only extrapolates the appropriate scale range for the correlation study between street view images and urban vitality in Shenzhen, but also provides a technical framework of scale selection for other cities in evaluating the impact of street view images on urban vitality, emphasizing the necessity for analytical scale tests before correlation studies.

2. Materials and Methods

2.1. Study Area

In this study, Shenzhen is chosen as the study area. Located in southern Guangdong Province and close to Hong Kong, Shenzhen covers nine administrative districts, namely, Futian District, Luohu District, Nanshan District, Baoan District, Guangming District, Longhua District, Longgang District, and Pingshan District, as well as one functional area, Dapeng New District, with a total area of 1997 km^2 and a resident population of 17.79 million. Shenzhen contains a variety of functional areas, such as modern business districts, high-tech industrial parks, and advanced manufacturing bases. There are obvious spatial differences in the socioeconomic development and population distribution of each area, presenting a wide variety of built environments.

2.2. Data Collection

We use Weibo check-in records to indicate urban vitality. These records provide information such as the geographic location of users when they perform check-ins on the Sina Weibo (https://open.weibo.com/, accessed on 22 November 2022). First, we crawl Weibo check-in records for Shenzhen from 21 August 2022 to 9 November 2022 and delete records with location names that do not match the corresponding latitude and longitude. This process led to obtaining a total of 101,299 valid Weibo check-in records. Second, we divide Shenzhen into grids of different sizes and count the frequency of Weibo check-in records within each grid. Given that the sampling interval of street view images is 100 m, the minimum grid scale size is set to 100 m. Since the human perception threshold for the environment is 1000 m [22], we choose 1000 m as the maximum scale. In total, 19 sizes of different grids are chosen, each in increments of 50 m (100 m, 150 m, 200 m, . . ., 950 m, 1000 m). The data distributions of Weibo check-in records within the above grids are not even and most of them feature low-value data. There will be problems such as convergence, bias, and overfitting of the training model if all data are used, which will make the regression results have a large deviation. To avoid the influence of extreme values and retain the overall characteristics of the distribution, we use the quantile sampling method to address data imbalances. The quartile method for data sampling of Weibo check-in records is applied at different scales in the subsequent study. For example, at a 500 m scale, all data were divided into four parts in order, and their minimum, first quartile (Q1), second quartile (Q2), third quartile (Q3), and maximum values were 0, 60, 125, 268, and 1661. The mean is 215, and the number of upper margins is 574. The rest of the dots are the discrete dots which are beyond the number of upper margins (Figure 1). Our experimental data were constructed by sampling from the street view images corresponding to each Weibo check-in frequency value (Figure 2a).

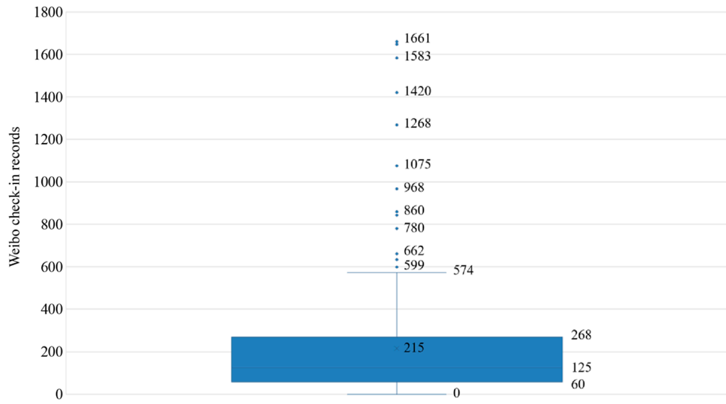


Figure 1. Boxplot of distribution of Weibo check-in records for 500 m grid. The blue dots reflect Weibo check-in records. The black lines represent the graphs vertical axis and its scale extension.

We use Baidu street view images to assess the built environment. The data acquisition process includes three steps. First, we extract one sampling point every 100 m along four categories of roadways: urban expressways, main roads, secondary roads, and side roads, in Shenzhen. A total of 246,350 sampling points are obtained. Second, we use the Baidu Map Open Platform (https://lbsyun.baidu.com/, accessed on 22 November 2022) to crawl Baidu street view images and obtain images captured at 0°, 90°, 180°, and 270° angles for each sampling point. This process leads to a total of 985,400 street view images to simulate the visual perception of pedestrians in the four directions. Finally, to characterize the built environment, all street view images we use have to contain building façade information, so we remove the street view images that do not contain building façade information. We use 10,000 manually labeled building images containing building façade information and 10,000 non-building images not containing building façade information as the training set and adopt the ResNet18 network for migration learning (an 18-layer convolutional neural network commonly used for image classification, with an accuracy greater than 95%) to classify building and non-building images from the acquired street view images. A total of 311,597 building façade-dominated street view images from Shenzhen (Figure 2b) are finally acquired in the subsequent scale effect study.

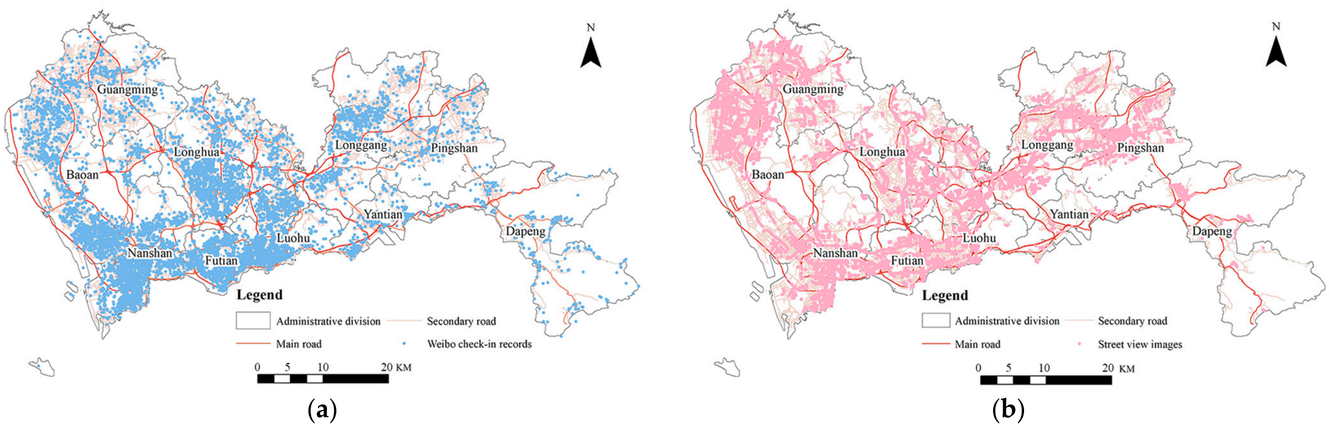


Figure 2. (**a**) Distribution of Weibo check-in records; (**b**) distribution of street view images.

2.3. Methodology

2.3.1. Visual Geometry Group (VGG) Models

We use visual geometry group (VGG) models to perform regression analysis between street view images and urban vitality. When street view images are processed, researchers often encounter challenges such as complex buildings, object occlusion, image distortion, brightness variation, and scale discomfort. It is difficult for traditional image processing methods to extract built environment features accurately [3]. VGG has good performance in dealing with street view images because of its deep structure, small convolution kernel, relocatable learning, and modular design [23,24]. Previous urban studies have confirmed that VGG has significant advantages in evaluating buildings, streets, and environments [25]. Therefore, we use VGG to learn features of street view images. Previous studies have mostly used VGG to obtain image classification results [26–28]. To explore the correlation between street view images and urban vitality, a VGG model is used to extract the feature tensor of the built environment with spatial connectivity. The VGG's fully connected layer is fine-tuned to determine the correlation between street view images and urban vitality. It is adapted to serve as the output layer for regression, which typically provides the probability distribution of the categories, enabling us to employ regression analysis of the street view images and urban vitality.

We design a VGG19-based regression neural network (a VGG model containing a 19-layer network) to perform the regression task and output a continuous value. A forward propagation function is used to pass street view images to each layer in the model step by step and finally output a regression prediction. The specific steps are divided into three stages: convolution, pooling, and full connectivity (Figure 3). First, we utilize VGG19, pretrained with the ImageNet dataset, for migration learning and use five sets of convolution layers to extract high-level features of the image. The initial two sets consist of two 3 × 3 convolution layers, and the last three sets consist of four 3 × 3 convolution layers. Each set of convolution layers is followed by a 2 × 2 max pooling layer (to reduce the feature size) and a Rectified Linear Unit (ReLU) activation function (to increase the nonlinear expressiveness of the model). Second, we introduce an adaptive average pooling layer to resize the extracted features (with 512 channels) to a fixed size (5 × 8). Finally, we add two fully connected layers to facilitate regression. To reduce the model complexity, we apply a flatten operation in the first fully connected layer to convert the image features into a one-dimensional vector with a length of 512 × 5 × 8 (a total of 20,480 features) and reduce it to 256 features (the number of neurons in the hidden layer). The second fully connected layer is used for the final regression output. The compressed 256 features are mapped to a single output value. Based on this model, the task of the regression analysis of street view images and urban vitality can eventually be realized.

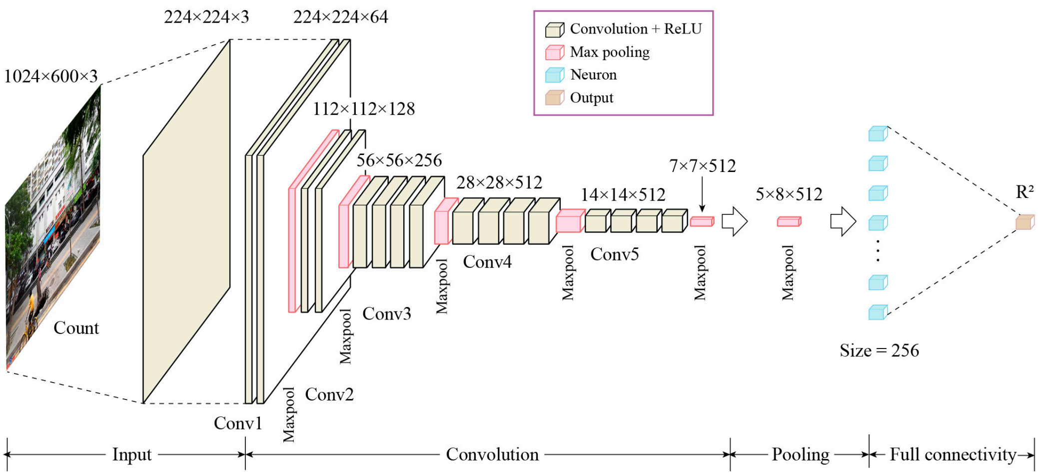


Figure 3. Structure of customized VGG19 regression.

Different grid sizes correspond to different Weibo check-in records; this leads to the same street view image being connected to different urban vitality values. To characterize the scale effect, we should train the model at each scale (divided into 19 grids). By comparing the goodness of fit of the regression model at each scale, we can obtain the law of correlation between street view images and urban vitality with scale changes, and then find the optimal scale. The model is trained with street view images and normalized labels (Weibo check-in records).

2.3.2. Loss Function

The loss function can be used to quantify the difference between the predicted values and the true values to evaluate the model performance during training. By minimizing the loss, we can find the optimal values of the model parameters to improve the prediction accuracy. Moreover, the loss gradient indicates the learning direction of the model, enabling parameter adjustment to reduce the error. The Mean Squared Error (*MSE*), which represents the average of the squares of the differences between the predicted and true values, is chosen as the loss function. *Loss* can be calculated as shown in Equation (1):

$$Loss = MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (1)$$

Here, y_i is the true value of the i-th sample, and $\hat{y}_i$ is the corresponding model prediction.

In VGG19 regression, the loss function is calculated from the training and validation sets. At each scale, we divide 200 street view images corresponding to the frequency of Weibo check-in records obtained via the quartile method into a training set, a validation set, and a test set. The training set consists of 150 images, the validation set consists of 25 images, and the test set consists of 25 images. Through multiple training cycles, the loss function is continuously updated after each iteration. Through evaluating the model performance with the validation set, we can identify and select the model with the best performance for regression analysis.

2.3.3. Goodness of Fit

Since previous studies have discussed the correlation between street view images and urban vitality, this paper does not use p-values for correlation tests, but calculates goodness of fit. The coefficient of determination (R^2) of the model with the test set is selected to measure the goodness of fit. This metric reflects the percentage of dependent variable variation that can be explained by the fitted model, which indicates how much of the variance of the dependent variable can be predicted from the independent variable. The R^2 value ranges from 0 to 1, and the closer to 1 this value is, the stronger the model's ability to explain variations in the data. We use spatial connections to enable all street view images to obtain the urban vitality values corresponding to the grid in which they are located and apply regressions to the two. R^2 is calculated in Equation (2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \quad (2)$$

where y_i is the true value of the i-th sample, $\hat{y}_i$ is the corresponding predicted value, and $\overline{y}$ is the mean of the true value.

3. Results

3.1. Spatial Distribution of Urban Vitality

The spatial distribution of urban vitality in Shenzhen is relatively consistent across different scales, following a pattern of 'high in the south and low in the north, high in the west and low in the east' (Figure 4). Specifically, areas with high vitality are more concentrated in Futian District, Nanshan District, and Luohu District in the southwest. In contrast, Baoan and Guangming districts in the northwest, as well as Pingshan and Dapeng New Districts in the east, mostly exhibit low urban vitality.

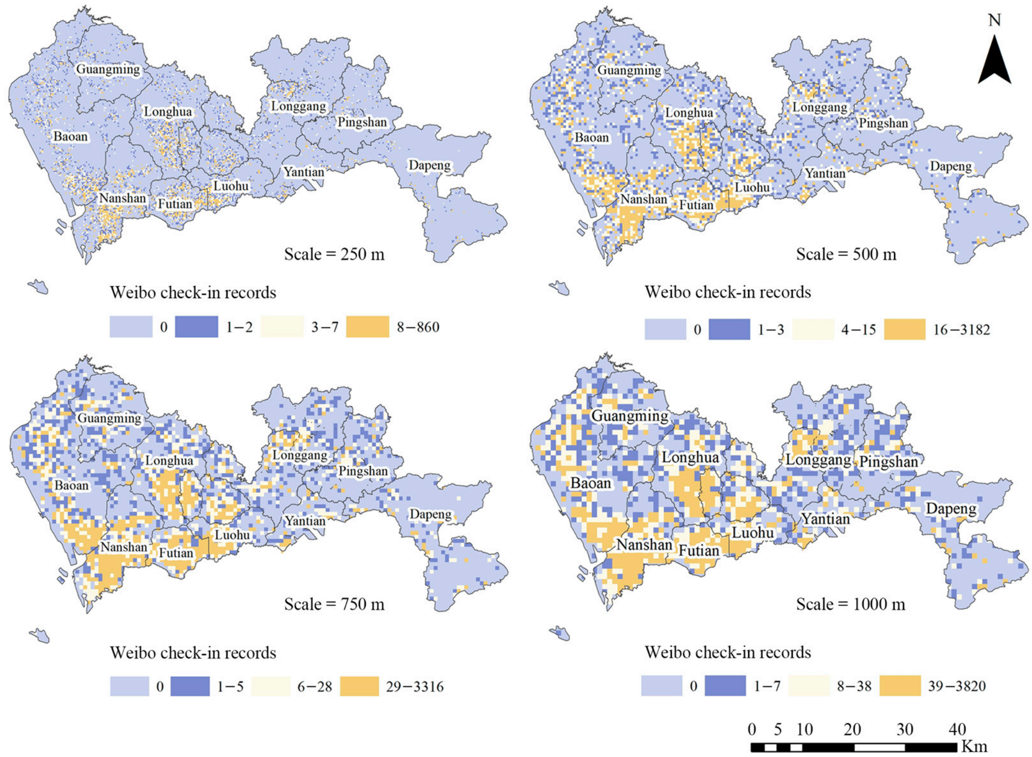


Figure 4. Distribution of urban vitality at multiple scales.

The spatial heterogeneity of urban vitality is influenced by the functional positioning of the area, which affects the design of the built environment. Futian District, Nanshan District, and Luohu District, which are highly vital, serve as hubs for economic headquarters, modern service industries, and technological innovation in Shenzhen. The overall built environment consists of high-rise buildings, shops along the street, and residential buildings. The Baoan District and Guangming District, which have low vitality levels, focus on the development of the advanced manufacturing industry as well as the air-side economy, and the built environment is mostly low-rise factory buildings. Pingshan District and Dapeng New District focus on advanced manufacturing, medical care, and recreation as well as tourism and holiday industries. The built environment includes low-rise factory buildings, small high-rise industrial parks, and scenic resorts. In summary, urban vitality is closely related to built environment features.

There are local differences in the spatial distributions of urban vitality across different scales. We find that as the scale increases, the area of high urban vitality in each district in Shenzhen increases. And the area of high urban vitality at a large scale is much larger than at a small scale. It is due to the fact that the larger the grid is, the larger the spatial division unit is, and the more Weibo check-in records point contained in each grid are accumulated.

3.2. Multi-Scale Correlation Between Street View Images and Urban Vitality

We evaluate the model by using the loss and R^2 value obtained using the training and validation sets. The model performance is optimized by continuously adjusting the loss (Figure 5a). As the number of epochs increases, the R^2 value tends to increase, and the explanatory power of the model significantly increases (Figure 5b). After 100 epochs, we all select the best model at each scale (Figure 5 shows the *Loss* and R^2 of the model for the 250 m, 500 m, 750 m, and 1000 m grids.). Taking the 500 m scale as an example, we select the model that performs best in the validation set, with a loss of 0.176 and an R^2 value of 0.836. The regression fit is high because the built environment features reflected by street view images can explain 83.6% of the changes in urban vitality.

We explore the scale effect of street view images and urban vitality by training a model to examine changes in R^2 across different scales. In general, as the scale increases, the loss value of the model first decreases and then increases, and the R^2 value first increases and then decreases. At the 500 m scale, the model has the smallest loss (0.176) and the largest R^2 value (0.836). The accuracy is 97.4%, indicating the best fitting effect (Table 1). This finding means that the correlation between street view images and urban vitality is high at the 500 m scale. Moreover, at scales less than 250 m and greater than 800 m, R^2 is less than 0.7. The model explains less than 70% of the changes in urban vitality, which indicates that the fitting effect is relatively low. The correlation between street view images and urban vitality is low.

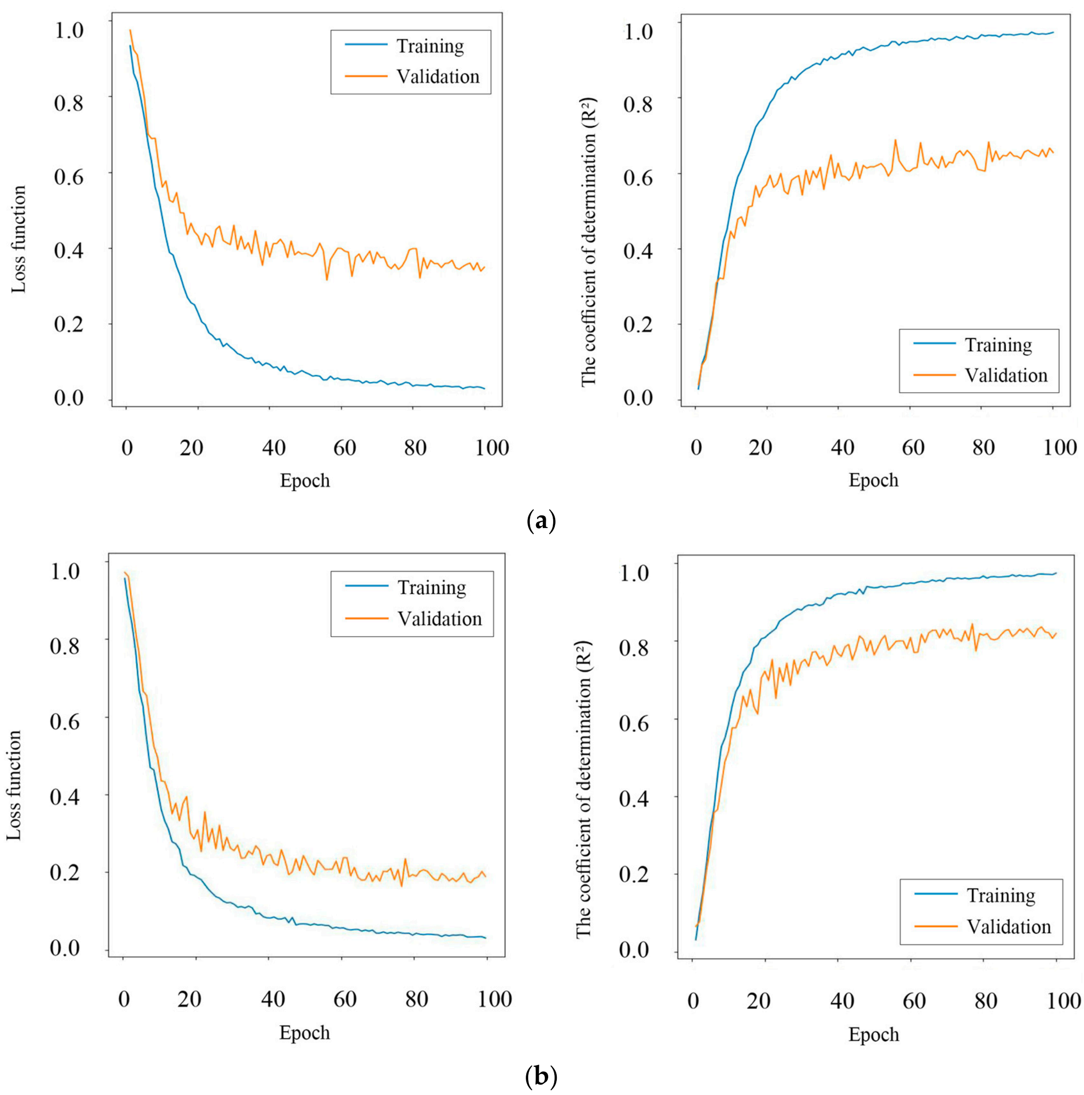
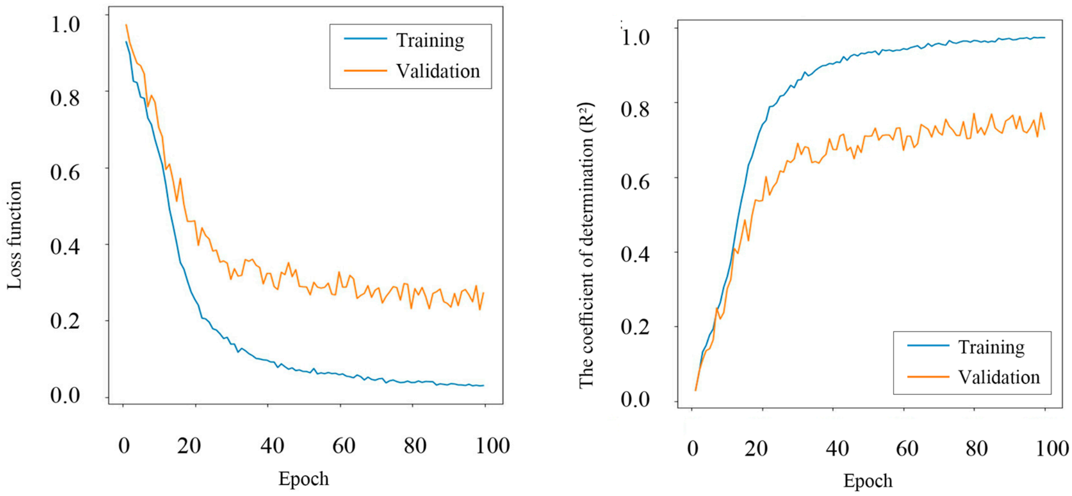
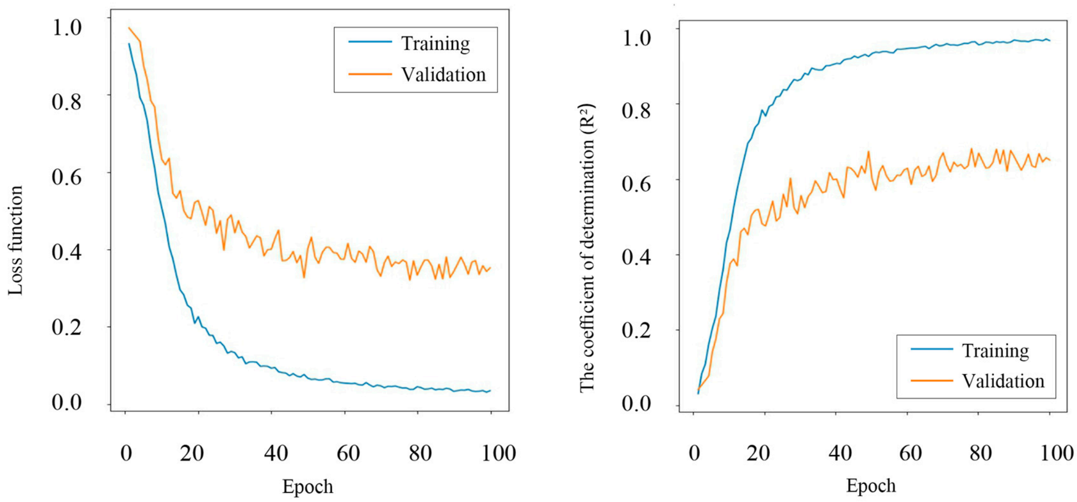


Figure 5. *Cont.*



(**c**)



(**d**)

Figure 5. Changes in *Loss* function and changes in R^2 values during training and validation on different scales: (**a**) 250 m scale; (**b**) 500 m scale; (**c**) 750 m scale; (**d**) 1000 m scale. In these figures, the horizontal coordinate "epoch" represents the number of iterations, meaning the times the model has traversed over the entire training dataset. It is the basis for the best performing model.

Table 1. Regression results at multiple scales.

Scale (m)	*Loss*	R^2	Scale (m)	*Loss*	R^2
100	0.438	0.567	600	0.326	0.708
150	0.486	0.537	650	0.270	0.762
200	0.354	0.654	700	0.237	0.727
250	0.383	0.642	750	0.274	0.726
300	0.287	0.741	800	0.370	0.684
350	0.210	0.789	850	0.371	0.640
400	0.244	0.759	900	0.318	0.671
450	0.252	0.763	950	0.442	0.620
500	0.176	0.836	1000	0.407	0.603
550	0.290	0.746			

3.3. Findings for the Scale Effect

To determine the appropriate scale of street view images in this type of study, we constructed scatter plots with scale on the horizontal axis and R^2 on the vertical axis, for different spatial scales. The results show that as the spatial scale increases, the intensity of the correlation between street view images and urban vitality presents a numerical distribution characterized with an increase and then a decrease, which is symmetrical and has a single peak. To quantify the scale effect, we need to choose a suitable function for fitting. The Gaussian function is the probability density function of the normal distribution, which can describe the probability distribution of random variables around the mean value. The distribution trend of the correlation between street view images and urban vitality is consistent with the law of change of the Gaussian function curve. Therefore, the Gaussian function is a suitable choice for this quantitative analysis.

We apply a nonlinear least squares method to fit these scatter points to the regression curve (Figure 6), and the fitting function is shown in Equation (3). The independent variable is the scale of the grids (19 grid sizes at 50 m intervals from 100 m to 1000 m), and the dependent variable is the degree of correlation between street view images and urban vitality at each scale. We find that Amplitude (a) is 0.773, Mean (μ) is 552.854, and Standard Deviation (σ) is 574.080. By calculation, the R^2 value of the curve fitting is 0.783, and the residual sum of squares (RSS) is 0.024. The fit is satisfactory, indicating that the scale effect of correlation between street view images and urban vitality largely conforms to the Gaussian function.

$$f(x) = 0.773 \cdot exp(-\frac{(x - 552.854)^2}{2 \cdot 574.080^2}) \tag{3}$$

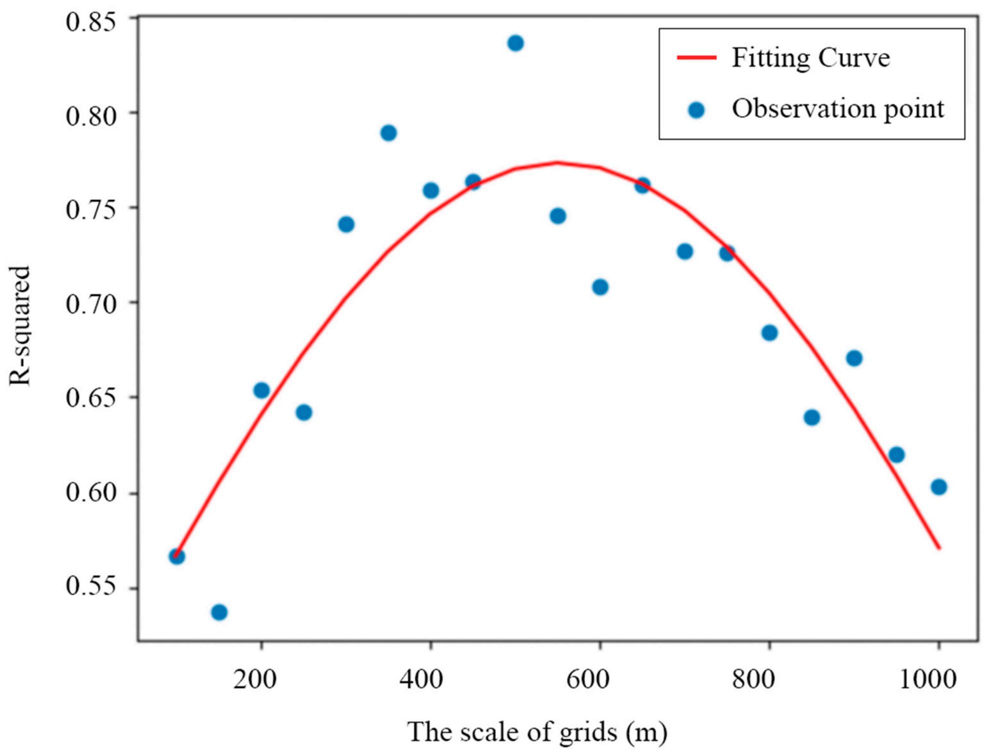


Figure 6. Gaussian function curve fitting plots for each grid scale.

4. Discussion

R^2 is statistically significant only within the appropriate scale range, suggesting that there is a scale effect of urban vitality and the built environment from an eye-level perspective. Built environment features can be assessed from both an overhead-view perspective and an eye-level perspective [29]. Previous studies have focused mainly on the scale effect of the built environment and urban vitality from an overhead-view perspective. Among overhead-view built environment features, the impact of land function has no obvious variation at different scales [30], whereas there are obvious scale effects of road form and architectural morphology on urban vitality. However, in previous studies of the correlation between urban vitality and the built environment from an eye-level perspective, relying on results obtained at a single scale may limit practical applicability [8]. Therefore, it is necessary to determine the scale effect of street view images and urban vitality. In this study, the analysis at multiple scales revealed that the correlation was significant within the appropriate scale range and the significance decreased beyond the appropriate scale range. This not only illustrates that the impact of the built environment on urban vitality from an eye-level perspective cannot be ignored, but also shows that the selection of the analysis scale affects the correlation of street view images and urban vitality.

The scale effect of urban vitality and built environment features assessed by street view images is different from the results of previous studies. The general conclusion of previous studies is that the correlation between the built environment and urban vitality shows a scale effect of continuous decay with increasing scale [18,20,31]. This difference may stem from previous studies often relying on intersection density, building coverage, and other indicators to describe built environment features from the overhead-view perspective [18,20], whereas this study employs street view images to assess the built environment from the eye-level perspective. The varying perspectives on assessing the built environment lead to different scale effect results. By comparing the results of the scale effect of two perspectives, it can be extrapolated that there is an appropriate scale for street view images.

The existence of appropriate scales for street view images is related to the laws of human perception. Street view images are perceptual data that simulate human visual observations of the built environment. However, human perception is limited at a smaller scale, overlooking discontinuities and roughness in the built environment [32]. It is difficult for the human eye to capture the differences in built environment features at smaller scales, leading to unreliable correlation analysis results when street view images are used. When the scale increases, the differences in built environment features are more obvious and can be perceived by humans. Then, street view images can be used to effectively assess the built environment and obtain reliable results. Owing to the limited range of human vision, humans are unable to perceive differences in the environment when they are beyond the range of human vision. Therefore, the use of street view images to assess built environments may not be effective when the scale extends beyond the typical range of human vision.

The scale effect of correlation between street view images and urban vitality conforms to the Gaussian function, indicating that there exists a threshold range of applicable scale for the correlation analysis. The finding is of general significance for other cities or regions. It reveals that researchers should test the applicable scale before conduct a correlation study between street view images and urban vitality to ensure the accuracy of the research results. This paper takes Shenzhen as the empirical research area; the result shows that the applicable scale is about 500 m. It is closely related to the cluster development pattern of Shenzhen's urban construction. Shenzhen's urban construction is based on a 500 m grid as a group unit. At this planning scale, unifying the building façade style with street functional attributes can enhance city visual continuity and spatial recognition, to create a more vibrant and distinctive neighborhood appearance. It also provides a reference for other cities to unify the size of zoning planning units for street façade design.

5. Conclusions

In this study, we determine different grid sizes to characterize scale changes, extract street view image features based on VGG19 to assess the built environment, determine the frequency of Weibo check-in records to reflect urban vitality, and explore the correlation between these variables across multiple scales. The results of this study show that as the spatial scale increases, the correlation between street view images and urban vitality shows a pattern of first increasing and then decreasing. The correlation indices are basically consistent with the Gaussian function. Unlike the results of the scale effect in previous studies, the correlation between overhead-view built environment indicators and urban vitality has continued to decrease. This finding indicates that there is an appropriate scale for street view images and that exceeding the availability scale range will lead to a decrease in the reliability of the results.

This study enriches the research on the correlation between street view images and urban vitality by explaining why previous studies obtained different results at different scales. Previous studies relying on results obtained solely at a single scale have limited their practical applicability, so it is essential to pay attention to the appropriate scale of street view images. Although recent studies in this area have identified optimal scales, these findings may not be universally applicable. Our study provides a methodological framework for selecting analytical scales when street view images are used. By detecting the scale effect via a Gaussian function, the range of appropriate scales for street view images can be popularized.

Taking Shenzhen as an example, this study reflects an attempt to investigate the appropriate scale of street view images for urban vitality assessment. Street view images can be used to assess the impact of the built environment on resident behavior, socioeconomic development, and thermal comfort. Future studies should include calibrating the appropriate scale of street view images by conducting similar studies with more indicators in more cities.

Author Contributions: Conceptualization, J.Y.; methodology, J.Y.; software, R.Z. and Y.F.; validation, R.C., R.Z. and X.L.; formal analysis, R.C.; investigation, R.C.; resources, J.Y. and Y.F.; data curation, R.Z. and Y.F.; writing—original draft preparation, R.C. and R.Z.; writing—review and editing, X.L.; visualization, R.C.; supervision, X.L.; project administration, J.Y. and X.L.; funding acquisition, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Shenzhen Stability Support Program under Grant [20220810221952002].

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. He, J.; Zhang, J.; Yao, Y.; Li, X. Extracting Human Perceptions from Street View Images for Better Assessing Urban Renewal Potential. *Cities* **2023**, *134*, 104189. [CrossRef]
2. Tang, J.; Long, Y. Measuring Visual Quality of Street Space and Its Temporal Variation: Methodology and Its Application in the Hutong Area in Beijing. *Landsc. Urban Plan.* **2019**, *191*, 103436. [CrossRef]
3. Zhang, F.; Salazar-Miranda, A.; Duarte, F.; Vale, L.; Hack, G.; Chen, M.; Liu, Y.; Batty, M.; Ratti, C. Urban Visual Intelligence: Studying Cities with Artificial Intelligence and Street-Level Imagery. *Ann. Am. Assoc. Geogr.* **2024**, *114*, 876–897. [CrossRef]

4. Wu, C.; Peng, N.; Ma, X.; Li, S.; Rao, J. Assessing Multiscale Visual Appearance Characteristics of Neighbourhoods Using Geographically Weighted Principal Component Analysis in Shenzhen, China. *Comput. Environ. Urban Syst.* **2020**, *84*, 101547. [CrossRef]
5. Kwan, M.-P.; Weber, J. Scale and Accessibility: Implications for the Analysis of Land Use–Travel Interaction. *Appl. Geogr.* **2008**, *28*, 110–123. [CrossRef]
6. Clark, A.; Scott, D. Understanding the Impact of the Modifiable Areal Unit Problem on the Relationship between Active Travel and the Built Environment. *Urban Stud.* **2014**, *51*, 284–299. [CrossRef]
7. Luo, P.; Yu, B.; Li, P.; Liang, P. Spatially Varying Impacts of the Built Environment on Physical Activity from a Human-Scale View: Using Street View Data. *Front. Environ. Sci.* **2022**, *10*, 1021081. [CrossRef]
8. Wu, C.; Ye, Y.; Gao, F.; Ye, X. Using Street View Images to Examine the Association between Human Perceptions of Locale and Urban Vitality in Shenzhen, China. *Sustain. Cities Soc.* **2023**, *88*, 104291. [CrossRef]
9. Deng, H.; Liu, K.; Feng, J. Understanding the Impact of Modifiable Areal Unit Problem on Urban Vitality and Its Built Environment Factors. *Geo-Spat. Inf. Sci.* **2024**, 1–17. [CrossRef]
10. Tao, Y.; Wang, Y.; Wang, X.; Tian, G.; Zhang, S. Measuring the Correlation between Human Activity Density and Streetscape Perceptions: An Analysis Based on Baidu Street View Images in Zhengzhou, China. *Land* **2022**, *11*, 400. [CrossRef]
11. Li, J.; Song, Z.; Wen, J.; Cai, C.; Tang, P. Exploring Nonlinear Relationship between Built Environment and Street Vitality Using Machine Learning. In Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia, CAADRIA 2024, Singapore, 20–26 April 2024; Gardner, N., Herr, C.M., Wang, L., Toshiki, H., Khan, S.A., Eds.; Caadria-Assoc Computer-Aided Architectural Design Research Asia: Hong Kong, China, 2024; Volume 2, pp. 375–384.
12. Li, K.; Lin, Y. Exploring the Correlation between Streetscape and Economic Vitality Using Machine Learning: A Case Study in the Old Urban District of Xuzhou, China. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 267. [CrossRef]
13. Xie, Y.; Zhang, J.; Li, Y.; Zhu, Z.; Deng, J.; Li, Z. Integrating Multi-Source Urban Data with Interpretable Machine Learning for Uncovering the Multidimensional Drivers of Urban Vitality. *Land* **2024**, *13*, 2028. [CrossRef]
14. Lan, F.; Gong, X.; Da, H.; Wen, H. How Do Population Inflow and Social Infrastructure Affect Urban Vitality? Evidence from 35 Large- and Medium-Sized Cities in China. *Cities* **2020**, *100*, 102454. [CrossRef]
15. Wu, C.; Ye, X.; Ren, F.; Du, Q. Check-in Behaviour and Spatio-Temporal Vibrancy: An Exploratory Analysis in Shenzhen, China. *Cities* **2018**, *77*, 104–116. [CrossRef]
16. Wu, J.; Ta, N.; Song, Y.; Lin, J.; Chai, Y. Urban Form Breeds Neighborhood Vibrancy: A Case Study Using a GPS-Based Activity Survey in Suburban Beijing. *Cities* **2018**, *74*, 100–108. [CrossRef]
17. Li, Q.; Cui, C.; Liu, F.; Wu, Q.; Run, Y.; Han, Z. Multidimensional Urban Vitality on Streets: Spatial Patterns and Influence Factor Identification Using Multisource Urban Data. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 2. [CrossRef]
18. Ge, Y.; He, Z.; Shang, K. Influence of the Built Environment on Pedestrians' Route Choice in Leisure Walking. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 384. [CrossRef]
19. Rui, J.; Xu, Y. Beyond Built Environment: Unveiling the Interplay of Streetscape Perceptions and Cycling Behavior. *Sustain. Cities Soc.* **2024**, *109*, 105525. [CrossRef]
20. Sarjala, S. Built Environment Determinants of Pedestrians' and Bicyclists' Route Choices on Commute Trips: Applying a New Grid-Based Method for Measuring the Built Environment along the Route. *J. Transp. Geogr.* **2019**, *78*, 56–69. [CrossRef]
21. Chen, J.; Wang, L.; Ma, L.; Fan, X. Quantifying the Scale Effect of the Relationship between Land Surface Temperature and Landscape Pattern. *Remote Sens.* **2023**, *15*, 2131. [CrossRef]
22. Alfonzo, M.A. To Walk or Not to Walk? The Hierarchy of Walking Needs. *Environ. Behav.* **2005**, *37*, 808–836. [CrossRef]
23. Iannelli, G.C.; Dell'Acqua, F. Extensive Exposure Mapping in Urban Areas through Deep Analysis of Street-Level Pictures for Floor Count Determination. *Urban Sci.* **2017**, *1*, 16. [CrossRef]
24. Huang, Z.; Wang, B.; Luo, S.; Wang, M.; Miao, J.; Jia, Q. Integrating Streetscape Images, Machine Learning, and Space Syntax to Enhance Walkability: A Case Study of Seongbuk District, Seoul. *Land* **2024**, *13*, 1591. [CrossRef]
25. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
26. Sharma, S.; Guleria, K. Deep Learning Models for Image Classification: Comparison and Applications. In Proceedings of the 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 28–29 April 2022; pp. 1733–1738.
27. Wang, Y.; Xiu, C. Spatial Quality Evaluation of Historical Blocks Based on Street View Image Data: A Case Study of the Fangcheng District. *Buildings* **2023**, *13*, 1612. [CrossRef]
28. Wu, Z.; Xu, K.; Li, Y.; Zhao, X.; Qian, Y. Application of an Integrated Model for Analyzing Street Greenery through Image Semantic Segmentation and Accessibility: A Case Study of Nanjing City. *Forests* **2024**, *15*, 561. [CrossRef]
29. Lu, Y.; Yang, Y.; Sun, G.; Gou, Z. Associations between Overhead-View and Eye-Level Urban Greenness and Cycling Behaviors. *Cities* **2019**, *88*, 10–18. [CrossRef]

30. Wang, X.; Zhang, Y.; Yu, D.; Qi, J.; Li, S. Investigating the Spatiotemporal Pattern of Urban Vibrancy and Its Determinants: Spatial Big Data Analyses in Beijing, China. *Land Use Policy* **2022**, *119*, 106162. [CrossRef]
31. Li, J.; Auchincloss, A.H.; Hirsch, J.A.; Melly, S.J.; Moore, K.A.; Peterson, A.; Sánchez, B.N. Exploring the Spatial Scale Effects of Built Environments on Transport Walking: Multi-Ethnic Study of Atherosclerosis. *Health Place* **2022**, *73*, 102722. [CrossRef] [PubMed]
32. Newburn, D.A.; Polsky, C.; Johnston, R.J.; Wang, H.; Ndebele, T. Modeling Multi-Scale Influences on Household Lawncare Decisions: Formal and Informal Neighborhood Conforming Effects on Fertilizer Use. *Landsc. Urban Plan.* **2023**, *240*, 104869. [CrossRef]